



OpenAIRE

Management výzkumných dat

Informační materiál

Únor 2016
Duben 2016 (Český překlad)

H2020-EINFRA-2014-1
Topic:e-Infrastructure for Open Access
Research & Innovation Action
Grant Agreement 643410

OBSAH

Zaměření publikace.....	2
Práva.....	2
Požadavky Horizontu 2020 & podpora OpenAIRE	3
Data, jejich management a využívání	3
Tvorba plánu managementu dat	6
Archivace dat a jejich zpřístupnění.....	7
Odpovědnost za RDM.....	8
Další užitečné materiály k managementu dat.....	10

ZAMĚŘENÍ PUBLIKACE

Cílem pilotního projektu otevřených výzkumných dat v programu Horizont 2020 je otevření výzkumných dat vybraných projektů. Tento stručný materiál se zaměřuje na všechny zúčastněné strany - vědce, projektové manažery a administrátory, NOADs, národní kontaktní body Evropské komise, pro které je téma otevřených výzkumných dat nové. Materiál ve své obecnosti nerozlišuje specifika jednotlivých vědních disciplín, jejich rozdílných workflow, standardů a datových formátů, ale slouží jako úvod do problematiky managementu výzkumných dat bez snahy o její úplné pokrytí. Pro zájemce o hlubší informace je určena poslední kapitola s přehledem odkazů na další literaturu.

PRÁVA

Toto dílo podléhá licenci [Creative Commons Uveďte původ 4.0 Mezinárodní License](https://creativecommons.org/licenses/by/4.0/).

Uloženo v repozitáři DSpace VŠB-TUO: <http://hdl.handle.net/10084/111436>

POŽADAVKY HORIZONTU 2020 & PODPORA OPENAIRE

Hlavní řešitel projektu z Horizontu 2020, který je zařazen do pilotního programu [Open Research Data Pilot](#) musí:

- Založit a udržovat aktuální plán managementu dat
- Uložit data do výzkumného datového archivu
- Zajistit třetím stranám volný přístup k datům a umožnit jejich vytěžování, využívání, reprodukování a šíření
- Vyjmenovat nástroje potřebné k využití dat za účelem ověření výzkumných výsledků, případně poskytnout přímo tyto nástroje

Projekt OpenAIRE podporuje pilotní projekt otevřených výzkumných dat, například informačními materiály o managementu výzkumných dat:

- Leták Open Research Data Pilot ([anglicky](#), [český překlad](#))
- [Webinář](#) Open Research Data Pilot (9. červen 2015, včetně slajdů)
- Webové stránky
 - [What is the Open Research Data Pilot?](#)
 - [How to create a Data Management Plan](#)
 - [How to select a data repository](#)
- [Často kladené dotazy](#)

DATA, JEJICH MANAGEMENT A VYUŽÍVÁNÍ

Odkřídlené úsloví říká, že Isaac Newton viděl dál díky tomu, že stál na ramenou obrů. Dnešní vědci mohou použít terabajty dat k řešení společenských a vědeckých otázek. Avšak efektivní využívání dat závisí na tom, jak efektivně jsou vytvořená data spravována a sdílena a zda jsou doplněna korektními metadaty a dokumentací.

„Management dat začíná od prvního dne“ : podstatnou součástí života vědce je plánování, jak nakládat s daty, jak je sbírat, používat, spravovat, ukládat, archivovat a šířit. O managementu dat více v oddíle 3 Tvorba managementu dat. Tato část přináší širší definici pojmů **výzkumná data** a **management (výzkumných) dat** (RDM), která je obecně použitelná ve všech vědních disciplínách.

Výzkumnými daty se myslí data ve formě faktů, pozorování, obrázků, výsledků počítačových programů, záznamů, měření a zkušeností, na jejichž základě je postavena vědecká teorie, hypotéza nebo jiný vědecký výsledek. Data mohou mít číselnou, textovou, obrazovou nebo hmatatelnou podobu. Data mohou být zpracovaná, očištěná nebo v surovém stavu a uložena v libovolném formátu a na libovolném médiu.

Management dat obnáší všechny procesy a aktivity spojené se správou dat v proběhu celého výzkumného cyklu a jejich zachováním pro výzkumné účely v budoucnosti.¹ V poslední době se management dat vnímá jako nedílná součást dobré výzkumné praxe (např. FTC², NWO³, RCUK⁴), která zaručuje možnost data studovat, replikovat, využívat a sdílet. Výzkum by měl alespoň na minimální úrovni plnit tyto požadavky. Pokud nemají být data volně dostupná, pak by alespoň měla být řádně spravována.

V pravidlech pilotního programu otevřených dat se Evropská komise zaměřuje na výzkumná data, dostupná v digitální podobě.⁵

Typy dat se liší v jednotlivých vědních oblastech a tím se mohou lišit i aktivity popsané v managementu výzkumných dat. V případě požadavku na opětovné využívání dat je však pro všechny oblasti shodná nutnost zajistit řádná metadata a dokumentaci k datům. Pilotní program otevřených dat vyžaduje navíc otevřený přístup k datům.



Obrázek 1 Životní cyklus výzkumných dat

Metadata jsou informace o výzkumných datech, které umožňují vyhledání dat, neboli se jedná o „standardizované strukturované informace vysvětlující účel, původ, časové a geografické údaje, údaje o autorovi, přístupových právech a podmínkách užití datového setu“⁶. Některé obory používají své metadatové standardy, viz adresář metadatových standardů tvořený Research Data Alliance⁷. V případě, že vědní disciplína nemá vlastní standard, doporučuje se použít metadatové schéma Dublin Core⁸ nebo schéma DataCite⁹. Jedná se o obecné standardy, běžně využívané vyhledávacími portály ke sklizení metadat z repozitářů a archivů. Správci repozitářů a datových archivů tyto standardy dobře znají¹⁰. Cílem vědců není tvorba vlastního metadatového standardu. Vlastním standardem by šli proti podstatě otevřené výměny dat a jejich interoperability s dalšími datovými zdroji.

K pochopení a užití výzkumných dat je potřeba poskytnout **dokumentaci** v podobě např. knihy kódů, laboratorního deníku, formuláře s informovaným souhlasem, případně **software** pro přístup a analýzu výzkumných dat. V dalších případech to může být syntaxe dotazů ve statistických analýzách, konfigurace nastavení měřicích zařízení aj.

V zásadě hlavní řešitel je odpovědný za tvorbu a archivaci celého souboru dat, metadat, dokumentace a softwaru. Jednoduše řečeno, vše, co je předmětem replikování a výzkumu, musí být součástí archivovaného souboru, včetně mezivýsledků nebo zpracovávaných verzí dat, pokud tyto verze odrážejí tvorbu nebo způsob analýzy ve výzkumném procesu. V praxi se stává, že není důvod archivovat průběžné verze dat například, když jejich nové generování je levnější, než samotná archivace. Takové rozhodnutí¹¹ ve výběru dat by mělo být dokumentováno spolu s postupem, jak data znovu získat.

Pilotní projekt otevřených výzkumných dat se zaměřuje na data, popisná metadata, dokumentaci, software a nástroje k pochopení a opětovnému využití dat. Jednoduše řečeno, pilotní projekt aplikuje požadavek na otevřený přístup ke všem datům, která jsou nezbytná k ověření výsledků uveřejněných ve vědeckých publikacích. Vědci mohou otevřeně sdílet také další data, když cítí, že tato data mohou mít nějakou hodnotu pro ostatní uživatele.

Vědci by měli specifikovat v plánu managementu dat (Data Management Plan, DMP), jaká data budou otevřeně sdílet. Co v tomto kontextu znamená výraz „otevřeně“? [Pravidla](#) pilotního projektu otevřených dat říkají: Otevřený přístup k výzkumným datům znamená právo na přístup a opětovné využití digitálních výzkumných dat. **Otevřeně přístupná výzkumná data jsou typicky volně přístupná pro uživatele a mohou být bez omezení vytěžována, využívána, reprodukována a šířena.**

TVORBA PLÁNU MANAGEMENTU DAT

Plán managementu dat (data management plan, DMP) slouží jako užitečný nástroj pro definování, jaká data, metadata a nástroje se budou v projektu používat a případně sdílet. „DMP je užitečný tam, kde vědci produkují data, zvláště pak tam, kde se na výzkumu podílí více partnerů, zemí atd.“¹²

V pilotním programu musí být DMP připraven během prvních šesti měsíců řešení projektu. Projektoví administrátoři a manažeři nebo vědci odpovědní za management výzkumných dat bývají odpovědnými osobami také za DMP. Vytvořením DMP v začátku projektu poskytovatelé podporují vytvoření nezbytných ujednání managementu dat v rané fázi projektu. Předchází se tak pozdnímu řešení problému, jak nakládat s daty v době, kdy je projekt

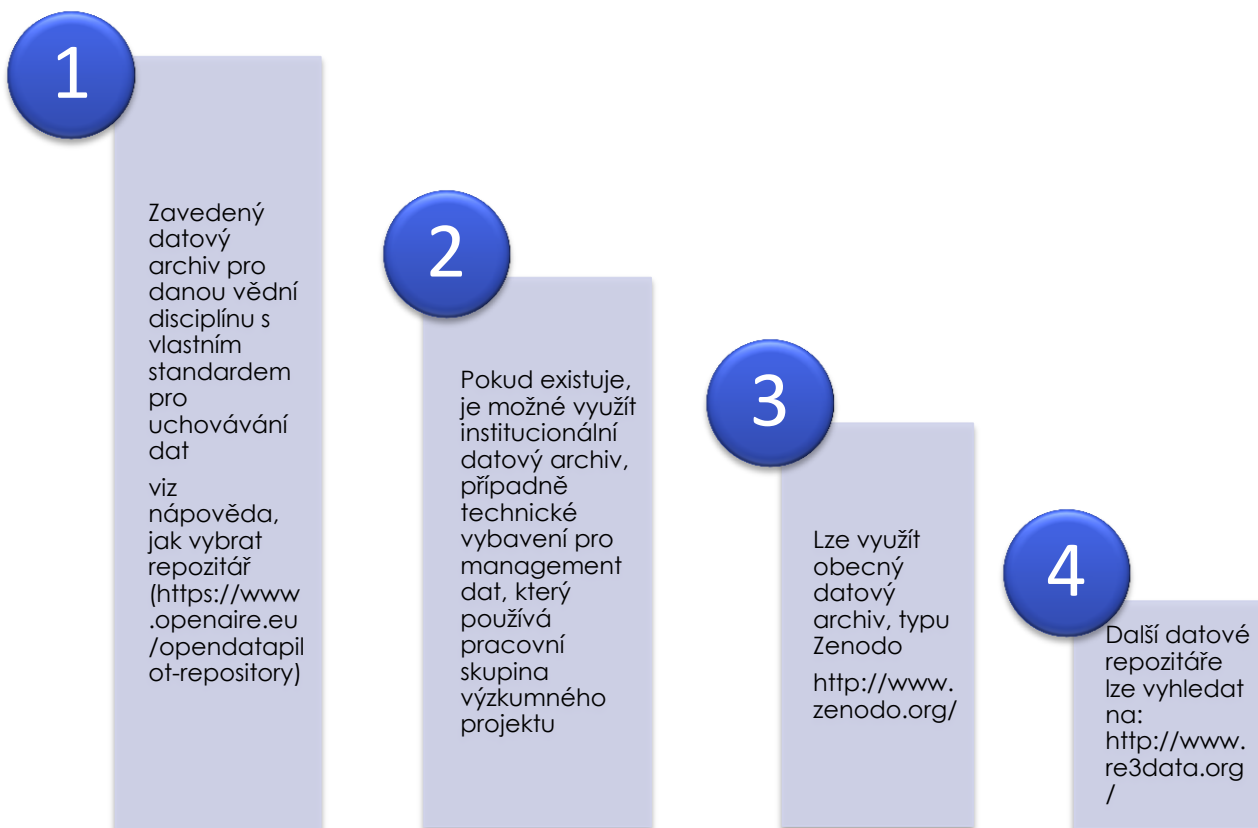
téměř před ukončením. K dispozici je **šablona DMP programu Horizont 2020** ([DCC's DMPonline tool](#)) a další informace na portálu OpenAIRE ([How to create a DMP Plan](#)). V Research Data Netherlands (RDNL) připravili videoklip [What, why and how of data management planning](#), který mimo jiné vysvětluje krok po kroku, jak vyplnit DMP. Dostupné jsou vyplněné vzory DMP (nejen pro Horizont 2020) [RDNL](#) a [DCC](#). [Videoklip](#) odpovídá na obvyklé obavy vědců z šíření jejich dat.

DMP je živým dokumentem, který se mění a zpřesňuje v průběhu řešení projektu. Může být aktualizován v průběhu a ve finální fázi projektu. Vedle formálních náležitostí by měl být DMP revidován v pravidelných intervalech a mělo by se kontrolovat, zda je dosaženo stanových cílů.

ARCHIVACE DAT A JEJICH ZPŘÍSTUPNĚNÍ

Smyslem pilotního projektu otevřených výzkumných dat v H2020 je zajistit přístupnost výzkumných dat s co nejméně restrikcemi při zachování ochrany neoprávněného zneužití citlivých dat. Projekty v pilotním programu v závěrečné fázi musí uložit soubor s daty, metadaty, dokumentací a nástroji do datového repozitáře, neboli archivu, který sbírá, uchovává a zpřístupňuje datové kolekce, související dokumentaci a metadata. Repozitáře obvykle používají termíny „preservation“ a „curation“ jako ekvivalenty k výrazům „archiving“ a „storage“. Dlouhodobá přístupnost vyžaduje odbornost a služby ke konverzi dat do nových formátů, k zajištění přidané hodnoty dat nebo např. k vytvoření nových možností, jak data vyhledávat.

Jestliže vědní obor používá své standardy a normy pro prezervaci dat, mohou být samozřejmě použity. Existuje více možností, podle nichž se lze rozhodnout, kam data uložit. Volba může být jednoduchá v případě, že pro danou vědní oblast funguje oborový datový archiv nebo infrastruktura, případně když řešitelská instituce buduje vlastní datový archiv. Preferované pořadí, kam data uložit:



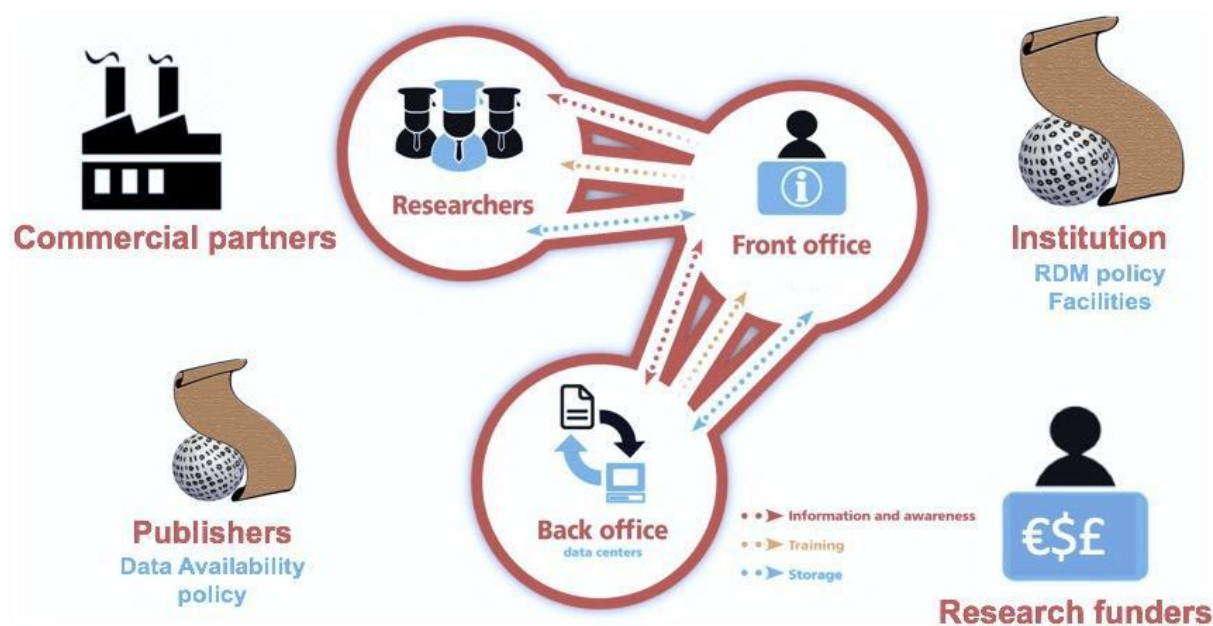


Dobrým kritériem pro výběr repozitáře může být jeho certifikace jako důvěryhodného digitálního repozitáře např. podle pravidel [Data Seal Approval](#).

Doporučuje se, aby řešitelé kontaktovali správce datového repozitáře ve chvíli tvorby DMP. Repozitáře mohou nabídnout pravidla a průvodce pro datové formáty a metadatové standardy, nebo rady, jak nakládat s citlivými údaji a licencováním dat. **Otevřené zpřístupnění dat zajistí vhodná forma licencování**¹³. Řada datových repozitářů akceptuje také ukládání publikací, čímž je zajištěno odkazování mezi publikacemi a použitými daty. Tím se také zvyšuje jejich viditelnost a potenciální opětovné využití.

ODPOVĚDNOST ZA RDM

Principy otevřené vědy podněcují různé zájmové skupiny, aby společně pracovaly na společenských cílech, které mají vzájemně shodné. Stejně tak RDM a plán managementu dat se týká rozličných účastníků.



Obrázek 2 Zúčastněné strany v RDM

- **Hlavní řešitel** – jednoznačně odpovědný za data a management dat
- **Vědci, vědeckí asistenti a/nebo datový manažer** – zabývající se běžně správou dat
- **Institucionální management** – navrhuje a podporuje datovou politiku, zvyšuje povědomí o datové politice
- **Institucionální oddělení složené z knihovny, IT oddělení a právního útvaru** – poskytuje data, nástroje, chráněné skladování a přístup k datům, řeší právní a etické otázky, citování dat, metadatové standardy, licencování, požadavky poskytovatelů, usiluje o zvýšení povědomí o datové politice
- **Poskytovatelé** – podporují vhodnou praxi ve zpřístupňování dat, investují do datové infrastruktury, zvyšují povědomí o datových politikách
- **Partneři projektu** – zástupci akademických, výzkumných i komerčních subjektů
- **Vydavatel** – podporují dostupnost dat a jejich propojení s publikovanými články; poskytují identifikátory k citování článků a souvisejících dat
- **Výzkumný datový repozitář** – uchovává dlouhodobě dat, poskytuje persistentní identifikaci a vyhledávací služby
- **Helpdesk** - NOAD (national open access desks) – poskytují odbornou radu v oblasti otevřeného přístupu k datům a publikacím obecně i v kontextu programu H2020

Každá ze zmíněných zúčastněných stran hraje v procesu otevřeného zpřístupňování dat určitou roli, proto je vhodné, aby vědci sdíleli svůj DMP se všemi stranami zapojenými do projektu.

DALŠÍ UŽITEČNÉ MATERIÁLY K MANAGEMENTU DAT

Management výzkumných dat se stává důležitým předmětem školení a kurzů vědeckých pracovníků, doktorandů a těch, kteří vědce podporují, a s nimiž spolupracují – knihovníků a odborníků na data. Níže zmíněné organizace nabízejí dobrý výchozí bod pro ty, kteří se chtějí dozvědět více o managementu výzkumných dat.

- **DCC** spravuje shrnutí [výukových materiálů](#) různých organizací. Uživatelé mohou sdílet navzájem přizpůsobené materiály v DCC a tím je dostat k širšímu okruhu zájemců
- **RDNL** nabízí kurz [Essentials 4 data supporters](#). Všechny materiály jsou licencovány jako CC-BY-SA 4.0. Všechny materiály mohou být volně užívané.
- **DataOne** nabízí powerpointové materiály [Data Management Education Modules](#). Materiály jsou licencované jako CC0 a uživatelé je mohou rozšiřovat a používat podle vlastních potřeb.

¹ Queensland University of Technology. (2013). [Management of Research data](#)

² The [Portuguese Foundation for Science and Technology](#)

³ The [Netherlands Organisation for Scientific Research](#)

⁴ Research Councils UK [data policies](#)

⁵ [Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020](#)

⁶ „standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection“ [RDNL course](#) Essentials 4 Data Support

⁷ Research Data Alliance (RDA) [Metadata Standards Directory Working Group](#)

⁸ [Dublin Core Metadata Initiative](#)

⁹ [DataCite Schema](#)

¹⁰ OpenAIRE průvodce pro správce repozitářů, [anglicky](#), [český překlad](#)

¹¹ Viz [DCC How to select what data to keep](#)

¹² “DMPs will be useful whenever researchers are creating or reusing data, especially where the research involves multiple partners, countries, etc.” [OpenAIRE Webinar Open Research Data Pilot in H2020](#) (Martin Donnelly, FOSTER/DCC). June 9, 2015.

¹³ Jak používat licence: <http://creativecommons.org/licenses/> nebo <http://opendefinition.org/guide/data/>.